

COKG-QA: Multi-hop Question Answering over COVID-19 Knowledge Graphs

Huifang Du¹, Zhongwen Le², Haofen Wang^{1†}, Yunwen Chen³, Jing Yu³

¹College of Design and Innovation, Tongji University, Shanghai 200092, China

²School of Computer Science, Fudan University, Shanghai 200433, China

³DataGrand Inc., Shanghai 201203, China

Keywords: COVID-19; Question answering; Knowledge graph; Knowledge embedding; Pre-trained model; Multi-hop KGQA

Citation: Du, H.F., et al.: COKG-QA: Multi-hop question answering over COVID-19 knowledge graphs. Data Intelligence 4(3), 471-492 (2022). doi: 10.1162/dint_a_00154

Data Citation: Du, H.F., et al.: COKG-QA: Multi-hop question answering over COVID-19 knowledge graphs. Data Intelligence 4(3), 2022. doi: <https://doi.org/10.57760/sciencedb.02062>

Received: Jan. 4, 2022; Revised: Apr. 10, 2022; Accepted: May 1, 2022

ABSTRACT

COVID-19 evolves rapidly and an enormous number of people worldwide desire instant access to COVID-19 information such as the overview, clinic knowledge, vaccine, prevention measures, and COVID-19 mutation. Question answering (QA) has become the mainstream interaction way for users to consume the ever-growing information by posing natural language questions. Therefore, it is urgent and necessary to develop a QA system to offer consulting services all the time to relieve the stress of health services. In particular, people increasingly pay more attention to complex multi-hop questions rather than simple ones during the lasting pandemic, but the existing COVID-19 QA systems fail to meet their complex information needs. In this paper, we introduce a novel multi-hop QA system called COKG-QA, which reasons over multiple relations over large-scale COVID-19 Knowledge Graphs to return answers given a question. In the field of question answering over knowledge graph, current methods usually represent entities and schemas based on some knowledge embedding models and represent questions using pre-trained models. While it is convenient to represent different knowledge (i.e., entities and questions) based on specified embeddings, an issue raises that these separate representations come from heterogeneous vector spaces. We align question embeddings with knowledge embeddings in a common semantic space by a simple but effective embedding projection mechanism. Furthermore, we propose combining entity embeddings with their corresponding

[†] Corresponding author: Haofen Wang (E-mail: carter.whfcarter@gmail.com; ORCID: 0000-0003-3018-3824).

schema embeddings which served as important prior knowledge, to help search for the correct answer entity of specified types. In addition, we derive a large multi-hop Chinese COVID-19 dataset (called COKG-DATA for remembering) for COKG-QA based on the linked knowledge graph OpenKG-COVID19 launched by OpenKG[®], including comprehensive and representative information about COVID-19. COKG-QA achieves quite competitive performance in the 1-hop and 2-hop data while obtaining the best result with significant improvements in the 3-hop. And it is more efficient to be used in the QA system for users. Moreover, the user study shows that the system not only provides accurate and interpretable answers but also is easy to use and comes with smart tips and suggestions.

1. INTRODUCTION

The serious situation of COVID-19 is ongoing. By January 16, 2022, more than 5.54 million people had died from the plague², raising increasing anxiety about health problems in individuals. The pandemic has severely affected people's lives, and people dramatically demand accurate, efficient, and instant access to epidemic information. However, large information about COVID-19 on various Web sites is not well organized and not specialized for the general public. Question Answering systems based on COVID-19 knowledge as a convenient interaction way are popular among more and more people. There are two existing paradigms for COVID-19 QA: Information Retrieval Question answering (IRQA) and Question Answering over Knowledge Graph (KGQA). The IRQA systems of COVID-19 are based on textual question-answer pairs [1, 2, 3, 4], getting answers by computing similarity between the asked question and questions/answers in the dataset. IRQA systems can naturally answer simple questions that people frequently ask. In contrast, KGQA methods over COVID-19 dataset [5, 6, 7, 8] give answers to complex questions covering multiple relations over structural KGs. Besides, KGQA techniques can reason for new knowledge in QA tasks.

On the other hand, the pandemic has been spreading for a long time until now, and people have some basic understanding of COVID-19. So people are no longer satisfied with asking simple questions, like "what are the clinical symptoms of patients with COVID-19?". They are more inclined to express complex multi-hop questions, such as the 2-hop question that "What are the related diseases having similar symptoms to COVID-19?" and the 3-hop question that "how to check the related diseases having similar symptoms to COVID-19?". So we choose to use multi-hop KGQA techniques to build COVID-19 QA system.

However, there are some limitations of existing KGQA techniques and current COVID-19 KGQA datasets. Existing methods [9, 10] often represent knowledge graph and questions by using separate models, raising issues that heterogeneous embeddings from different spaces should be fitted to a common space. Additionally, a schema that defines a useful, high-level structure of a KG has been neglected in the current multi-hop KGQA tasks [11]. Schema information as important prior knowledge can be helpful to search for correct entities of specified types. What's more, public COVID-19 KGs [12, 13, 14] suffer from knowledge sparsity especially the knowledge people would like to ask for daily, which will further affect the quality of

[®] <http://openkg.cn>, the largest Chinese open knowledge graph community pushing the development of public KGs, open-source tools, and best practices in vertical sectors in China.

downstream QA tasks. In this paper, we improve KGQA performance by proposing COKG-QA: multi-hop Question Answering over COVID-19 Knowledge Graphs. COKG-QA proposes some improvements in terms of these constraints mentioned above. The architecture of our system is illustrated in Figure 1, and the main contributions of our paper are as follows:

- 1) We introduce COKG-QA to demonstrate the importance of embedding projection mechanism and schema information in multi-hop KGQA task^②. More precisely, embeddings of entities, schema, and questions from different spaces are transferred into one common one by a projection method to align important features. Furthermore, entity embeddings are incorporated with its type embeddings to predict answers of specified types.
- 2) There rarely exist comprehensive KGQA datasets^③ managed for COVID-19 especially lacking multi-hop questions. Benefiting from OpenKG-COVID19 [15], we derive a large multi-hop Chinese COVID-19 KGQA dataset, COKG-DATA. It consists of abundant knowledge, which provides an important foundation for building a superior question answering system.
- 3) Experiments in the paper prove that COKG-QA is of high quality and also robust to further generalize to new knowledge. In order to facilitate people's demand for COVID-19 consulting services, we develop a user-friendly interactive application^④ based on COKG-QA. The application not only provides accurate and interpretable answers but also is easy to use and has functions of smart tips and recommendations.

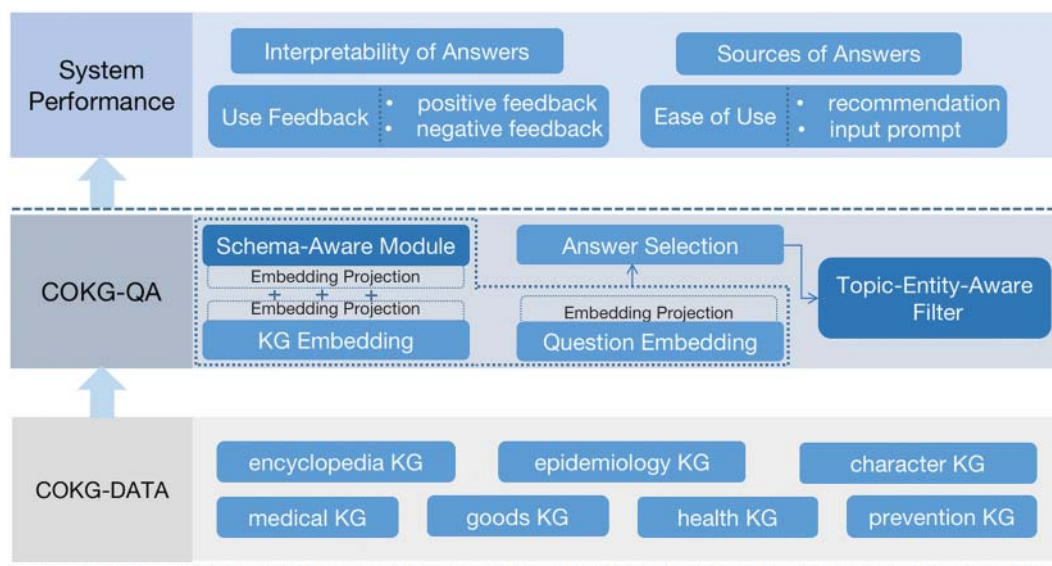


Figure 1. Architecture of the COKG-QA system.

^② <https://news.qq.com/zt2020/page/feiyan.htm#/>

^③ Our code and data are available at <https://github.com/mug2mag/COKG-QA> and <http://openkg.cn/dataset/cokg-data>

^④ You can access the system at <http://cokg-qa.openkg.cn/qa/>

2. RELATED WORK

At the moment of the epidemic, researchers have released some relevant datasets and built question answering systems based on natural language processing techniques to help people conveniently obtain information about COVID-19. We introduce these significant efforts and KGQA techniques in this section.

COVID-19 datasets and QA systems: Some useful KGs have been launched to advance COVID-19 research during the ongoing pandemic. However, the published COVID-19 KGs have limited data size and are more academically medical, which are not applicable for users' daily consulting needs. For example, the coronavirus Knowledge Graph [13] has 27 relations and limited entity types. Publications, case statistics, and molecular data are structured [12] to explore biomedical knowledge, such as specific genes, proteins, etc. KG-COVID-19 [14] also focuses on SARS-CoV-2 and COVID-19 related heterogeneous biomedical data to construct KGs. Based on the public KGs, some KGQA systems are developed. Template matching method [7] using Naive Bayes algorithm over a KG is adopted to establish a QA system of COVID-19. QA system like [16] employs a rule-based classifier for recognizing users' intentions and also adopts templates to parse natural questions of users. To make the framework not limited to predefined rules, some work like [17] introduces a relatively general framework based on the knowledge embedding method tranE [18]. Although these QA systems are developed for COVID-19, they fail to provide optimal performance for users' diverse questions.

KGQA: There are many state-of-the-art KGQA methods, and we briefly review these three types [19]: (1) logic-based methods; (2) path-based methods; (3) embedding-based Methods. Logic-based methods are widely discussed due to the advantages of high accuracy and strong interpretability. GQE (Graph Query Embedding) [20], Query2Box [21], BetaE [22] represent the query as a directed acyclic computational graph to generate logic form query embedding. Pathbased methods take the topic entity in the question to search along multiple triples of KG to find the answer entity or relation. To alleviate the issue that the search space of Path-Ranking Algorithm [23] is large, DeepPath [24] allows the path attributes to be controllable. Teacher-student network is adopted in NSM [25] to learn intermediate supervision signals. Some other works like [26, 27] regard KG reasoning as a sequential path decision process. Embedding based methods [11, 28] measure the similarity between question embeddings and candidate answer embeddings to get the right answer. For example, the state-of-the-art method EmbedKGQA represents questions by pre-trained model and represent knowledge graph embeddings by ComplEx [29], and select answer through the score function of ComplEx. Relational Graph Convolutional Networks (R-GCN) method [30] aggregates embeddings of specific multiple relations in KG to predict answers. Research that KGs incorporate text corpus based on embedding methods [9, 10, 31] also attract much attention.

3. TASK DEFINITION

To alleviate people's anxiety about health problems caused by the COVID-19 pandemic, we are determined to develop an effective KGQA system focusing on complex multi-hop questions. In addition, the functions of smart tips and recommendations make the QA system consumer friendly. Considering the questions tend to be asked daily, data derived from OpenKG-COVID19 will be curated elaborately and

finally formed into the multi-hop KGQA dataset, COKG-DATA. Moreover, we propose COKG-QA extending the state-of-the-art EmbedKGQA model with simple and efficient modifications, so that it can achieve superior and practical performance for the QA system. In the following sections, we first describe the extension in COKG-QA and the details of COKG-DATA. Based on the two modules, the performance of the KGQA system will be demonstrated finally.

4. COKG-QA

As mentioned in the related work, EmbedKGQA [11] is a good work considering multi-hop reasoning. We extend it in the following several aspects to achieve better performance in terms of accuracy and coverage in the context of COVID-19 question answering. We first give a brief introduction of EmbedKGQA and then describe our improvements in detail in the following subsections.

4.1 Preliminary

An instance triple in a KG can be represented as $\langle h, r, t \rangle$, where h represents the head entity and t represents the tail entity linked by relation r . Given a set of entities E and relations R , a Knowledge Graph G is a set of triples K such that $K \subseteq E \times R \times E$. KGQA task searches answer entity for a natural language question q including multi-hop relations over a KG. Inspired by EmbedKGQA, we also employ KG Embedding Module, Question Embedding Module, and Answer Selection Module in our method. In this paper, we extend EmbedKGQA over COKGDATA by adding Embedding Projection and Schema-Aware Module. In addition, we also add a Topic-Entity-Aware Filter at inference to predict answer entity only related to the topic entity in question. The architecture can be seen in Figure 2. Details are described as follows.

4.2 Embedding Projection

We regard embeddings generated by different models as heterogeneous. Like triples in instance level, $\langle s, h, r, s, t \rangle$ is a triple in schema level, where s, h represents the head type and s, t stands for the tail type linked by relation r . Schema embeddings of $s, h, s, t \in E'$ are also trained by ComplEx [29] method to enhance searching answer, but schema model and instance model are trained separately. What's more, question embedding is produced by pre-trained model RoBERTa [32] which leverages quite another technique paradigm. Therefore, these three embeddings are heterogeneous. Even though it helps to maintain their characteristics of schema, instance, and question by separate models, it is hard to model representations in the final KGQA model. Fully Connected (FC) linear layers like "firewalls" can maintain and project important features in transfer learning [33], especially when the source domain and target domain are quite different. Therefore, it is reasonable to project these embeddings before being transferred into one common space. We respectively define question embedding, entity embedding, schema embedding by

$$E_q = FC(e_q), \quad (1)$$

$$E_n = FC(e_n), \quad (2)$$

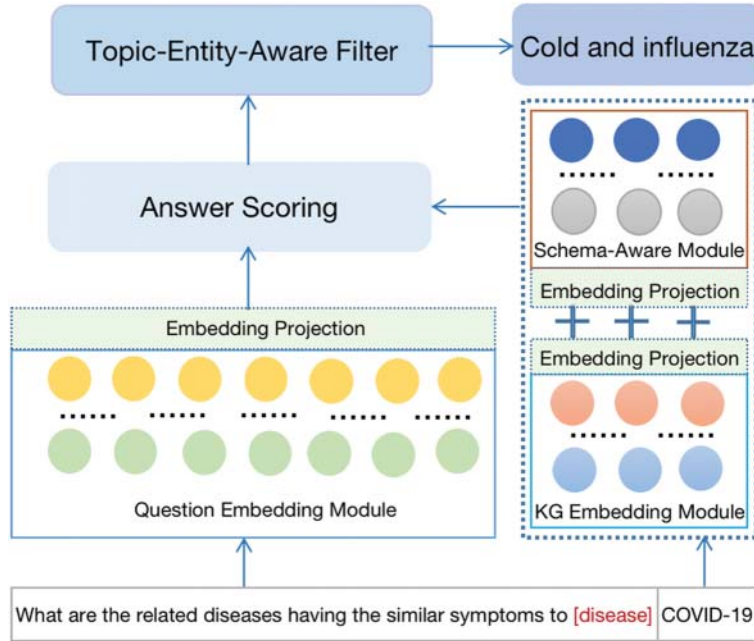


Figure 2. Overview of COKG-QA framework.

$$E_{s_n} = FC(e'_{s_n}), \quad (3)$$

where e_q is question embedding. And e_n is entity embedding trained by instance triples, while e'_{s_n} is entity type embedding produced by triples in schema level.

4.3 Schema-Aware Module

Existing KGQA methods only focus on instance facts in the KG, which ignores the well-constructed prior knowledge in the schema. The schema contains valuable structure information of a KG, which defines concepts and properties of these concepts. Entities in KG are linked to their corresponding concepts by entity types [34, 35]. We add Schema-Aware Module by combining entity embedding with corresponding entity type embedding which will be helpful to filter answer entities of specified types. This is good enough for the model to understand which type of the topic entity is and which type of the answer entity will be. Specifically, the topic entity representation in the question and the tail entity representation as the answer is constructed by adding the corresponding entity type embedding. Question representation embedded by using RoBERTa can't encode relation embedding in the schema level because there is no relation type label for question in a real application. But we concatenate entity type with the given question to imply that the question is relevant with a certain entity type like the input shown in Figure 2.

And the specific function is

$$\phi(E_h + E'_{s-h}, E_q, E_a + E'_{s-a}) > 0 \quad \forall a \in \sigma, \quad (4)$$

$$\phi(E_h + E'_{s-h}, E_q, E_{\hat{a}} + E'_{s-\hat{a}}) > 0 \quad \forall \hat{a} \in \sigma, \quad (5)$$

where ϕ is the ComplEx scoring function described in section 4.1 and E_h is the topic entity embedding and E'_{s-h} is its corresponding type embedding. E_a stands for the right answer entity and $E_{\hat{a}}$ means negative entity. $\sigma \in E$ is the set of answer entities. All these embeddings are all transferred by Embedding Projection.

4.4 Topic-Entity-Aware Filter

Because COKG-DATA we collect is very large, it's necessary to add a filter to get the topic entity⁹ related entities, including 1-hop, 2-hop, and 3-hop entities at inference like EmbedKGQA to predict more relevant answer entity. We first make a map between topic entities and its multi-hop entities with 3-hop number, and then we predict answers among the multi-hop entities based on the best-trained model.

5. COKG-DATA CONSTRUCTION

Existing COVID-19 QA systems fail to perform complex reasoning with a non-KG dataset. We organize COKG-DATA based on seven sub-KGs (i.e., encyclopedia, prevention, goods, medical, epidemiology, character) of OpenKG-COVID19 launched by OpenKG, which people are more prone to ask daily. COKG-DATA is a new challenging question-answer benchmark that contains single-hop questions and multi-hop questions concerning diseases, symptoms, drugs, etc. The overview of the selected graphs by COKG-DATA is depicted in Appendix A.1. With the large and diverse COKG-DATA, multi-hop KGQA is an appealing and useful task to satisfy people's complex query needs during the pandemic. We spend much time cleaning data based on OpenKG-COVID19 and collecting multi-hop questions. Details are shown in A.2.

5.1 Human Check

To make sure that the questions in COKG-DATA are natural and meaningful, we recruited four volunteers whose research fields are all Knowledge Graph and Question Answering to check the quality of the dataset. We got random samples in proportion to the number of the questions sorted by each relation defined in the cleaned OpenKG-COVID19. These four volunteers were asked to rate the sampled questions with three choices: 1 for Weird; 2 for Natural; 3 for Meaningful. We have optimized COKG-DATA four times by removing or modifying the weird question-answer pairs through the scoring process. The sampled number for the last turn is 4, 000, and the average score by volunteers is 2.8 demonstrating the high quality of COKG-DATA. The final statistics for each hop questions of COKG-DATA are shown in Table 1. COKG-DATA will keep up with OpenKG-COVID19 to update for more sufficient knowledge for users.

⁹ We are only concerned about questions with single topic entity in this paper.

Table 1. Statistics for COKG-DATA.

Dataset	Train	Dev	Test
COKG-DATA 1-hop	165,795	55,239	55,239
COKG-DATA 2-hop	48,147	16,049	16,049
COKG-DATA 3-hop	2,811	927	927

6. EXPERIMENTS

In this section, we first present the experimental setup, the COKG-QA results on COKGDAT, and then analyze answer errors.

6.1 Experimental Settings

We follow the same split proportion (i.e., 3:1:1) of train/validation/test for all datasets of 1-hop, 2-hop, and 3-hop questions. The number of each hop questions are summarized into Table 1. We choose batch size of 90, 64, 32 and corresponding learning rate $5e-5$, $2e-5$, $1e-6$ for training model across 2 NVIDIA RTX2080ti GPUs. Additionally, we set the patience number as 10 meaning that it will stop training when the accuracy score has decreased ten times and the maximum limitation epoch is 100. ComplEx embedding was obtained based on OpenKE[®] and the dimension of ComplEx embedding and question embedding in COKG-QA are all 400. Weight decay as a popular and necessary regularization technique was set as $1e-1$.

6.2 Baselines

We compare our model with two state-of-the-art models, including EmbedKGQA [11] and TransferNet [28]. Since EmbedKGQA reasons answer through link prediction which can alleviate the KG incompleteness problem and avoid the problem of uneven distribution of data, we take extensions over it in our implementations. TransferNet is an Effective method and competitive enough as a baseline which achieves best performance on public multi-hop datasets, such as MetaQA [41], WebQSP [42], and CompWebQ [43].

EmbedKGQA [11] regard multi-hop KGQA task as link prediction and search for answer entity based on question embedding and knowledge embeddings, which mitigates the problem of KG incompleteness and can predict answer in unlimited neighbors.

TransferNet [28] proposes a unified architecture for label and text data. In this framework, TransferNet calculates the relations corresponding to different positions of the question under attention mechanism at each step and further gets the answer entity.

[®] <http://openke.thunlp.org/>, an Open-source Framework for Knowledge Embedding.

6.3 Main Results

In Table 2, we compare EmbedKGQA and TransferNet with COKG-QA on our COKGDATA datasets. COKG-QA performs better than EmbedKGQA in all hop data, while TransferNet outperforms COKG-QA in 1-hop and 2-hop questions. But TransferNet obtains the lowest accuracy in the 3-hop questions. TransferNET attends to different parts of the question to search for the corresponding relation at each step, which makes it sensitive to both the quality and quantity of each-hop relations in the graph. Therefore, we assume that the small amount of 3-hop data of COKG-DATA causes the bad performance for TransferNET. However, EmbedKGQA and COKG-QA both regard the multi-hop KGQA task as link prediction which takes a multi-hop relation as a single relation in KG Embedding Module. For example, each relation of “complication||commonly used medicine||usage and dosage”, “medication||medication ingredient” and “precaution” is equally seen as a single relation to put in one triple. So COKG-QA avoids the problem of data imbalance which is very common in the real world and poses challenge to neural models. What’s more, TransferNET has a high complexity of computation and large memory storage problems because it computes the probability of an entity being activated as the answer entity for multi-times, which would also affect the inference speed.

Table 2. Results of COKG-DATA with improvements.

Model	1-hop	2-hop	3-hop
EmbedKGQA	73.19	80.70	88.59
TransferNet	99.58	96.36	11.50
COKG-QA	95.75	92.90	97.30

6.4 Ablation studies

Table 4 shows ablation studies of the effects of adding Schema-Aware Module, adding Embedding Projection and Topic-Entity-Aware Filter. We demonstrate the importance of each improvement by leveraging the same train set, validation set, test set, and hyperparameters. We briefly analyze the effect of each component in this section.

6.4.1 Effect of Embedding Projection

Since all the entities embedding are frozen during COKG-QA training as EmbedKGQA does, the features of entities embedding are quite different from question embedding. Besides, entity embeddings and type embeddings are also learned from different trained models. So it is necessary to bridge a projection to transform these important features in different vector spaces into a common vector space. The comparison results of projection (in COKG-QAep row) and without projection can be seen in Table 3. Although Embedding Projection does not provide as much improvement as Schema-Aware Module, the 2.61% absolute improvement in 1-hop questions and soft better performance in other questions demonstrates that Embedding Projection advances the capability of COKG-QA compared to EmbedKGQA.

Table 3. Results of COKG-DATA with improvements.

Model	1-hop	2-hop	3-hop
EmbedKGQA	73.19	80.70	88.59
COKG-QA _{ep}	75.80	81.31	90.40
COKG-QA _{sam}	77.54	82.64	88.56
COKG-QA _{teaf}	90.84	92.43	96.98
COKG-QA	95.75	92.90	97.30

Note: The results reported in this table are hits@1. The subscript of COKG-QA_{sam} is named by the first letter of each word of Schema-Aware Module.

COKG-QA_{ep} for adding Embedding Projection, and COKG-QA_{teaf} for Topic-Entity-Aware Filter.

6.4.2 Effect of Schema-Aware Module

We concatenate entity embedding to the corresponding entity type embedding to build a contextual KG embedding for COKG-QA. Furthermore, an ablation test was performed to evaluate the effect of the only Schema-Aware Module. The results listed in Table 3 marked by COKG-QA_{sam} show Schema-Aware Module leads to a better performance of an average increase by 1.82%, which indicates the effectiveness of enriching entity embedding by adding schema information.

6.4.3 Effect of Topic-Entity-Aware Filter

To select an answer entity in the range of the 3-hop neighborhoods of the topic entity, the filter could competitively deliver better inference results with more than a 10% increase, which further ensures to provide a robust QA system on COKG-DATA.

6.5 Answer analysis

To fully analyze the results of the experiments, we collect all wrong answers of the test set to try to find some useful reasons. Through observation, we find that the wrong samples containing digital numbers (in their digit or word form) account for 33.92%. And there are about 11.94% percentage entities including numbers in the selected sub-graphs, which is not a negligible data size. Numerical reasoning or discrete reasoning is a more challenging task [36] with only question-answer pairs supervision. Therefore, we experimented with two types of data, i.e, numerical question-answer pairs (inserted with numbers) and non-numerical question-answer pairs, to probe the impact of data types. We also tested their corresponding 2-hop, 3-hop questions. Table 4 shows the results for different types of datasets using our model.

Table 4. Performance of COKG-DATA by different data types.

Dataset	1-hop	2-hop	3-hop
COKG-DATA _{numeric}	49.19	62.87	76.38
COKG-DATA _{non-numeric}	85.41	81.46	81.46
COKG-QA _{no-teaf}	80.06	81.01	90.72

Numerical Data Analysis. COKG-QA over only numerical data reaches 49.19% hits@1 in the 1-hop data with a 30.87% absolute decrease compared to the model with all data (without Topic-Entity-Aware Filter condition). It highlights the fact that it is harder to model text with numbers. Besides, entities and relations distributions in the numerical dataset are also observed and show that the uneven distributions may be another key factor for the worse performance. The right histogram in Figure 3 gives entities and relations distributions of Numerical Data.

Non-numerical Data Analysis. As expected, non-numerical data with large samples is still hard to optimize, because non-numerical data accounts for the majority and the distribution of the non-numerical dataset is similar to all data. However, without the numerical problems, the experimental results of non-numerical data are better than of all data. The left histogram in Figure 3 presents the visualization of COKG-DATA distribution according to the first 30 multi-hop relations sorted by entity number. We can see that both numerical and non-numerical data have long-tail data problems, for which data augmentation to compensate [37] or enhancing the recall of long-tail entities [38] are directions that can be considered.

7. COKG-QA PERFORMANCE

The superior performance of COKG-QA illustrated by the extensive experiments above will promise an effective QA system. Therefore, we devise an interactive Web QA application based on COKG-QA for people. A friendly design of QA system can improve user experience [39, 40]. We discuss the considerations designed in the QA application in this section.

7.1 Interpretability of Answers

Unlike most KGQA systems giving direct answers, our system will explain the intermediate context for the multi-hop questions to make the answer for multi-hop questions interpretable. An answer will be inferred based on the best-trained model by computing ComplEx score. But the answer based on EmbedKGQA model is not understandable. For example, the answer to the 2-hop question “What are the types of drugs recommended for pediatric intracranial tumors” is “Chemical drugs, prescription drugs and medical insurance drug for work-related injury”, which would pose users a question like “what are the respective recommended drugs corresponding the drug types mentioned in the answer above?”. In other words, people not only want to achieve the final answer but also want to figure out what the intermediate results are. So we offer an interpretable answer “The recommended drug for pediatric intracranial tumors glycerol fructose injection is a chemical drug; the recommended drug for pediatric intracranial tumors piracetam glucose injection is a medical insurance work injury drug...”. The process for the interpretable response is as follows: (1) When the QA system gets a multi-hop question, the topic entity will be recognized first. (2) Subsequently, the not direct tail answer is obtained by ranking scores based on the question and the recognized head. (3) To get an interpretable final answer, we need to search out the intermediate relations and get intermediate entities. Questions and corresponding multi-hop relations having the same head and answer labeled in the dataset are filtered out. Furthermore, we select the interpretable answer corresponding

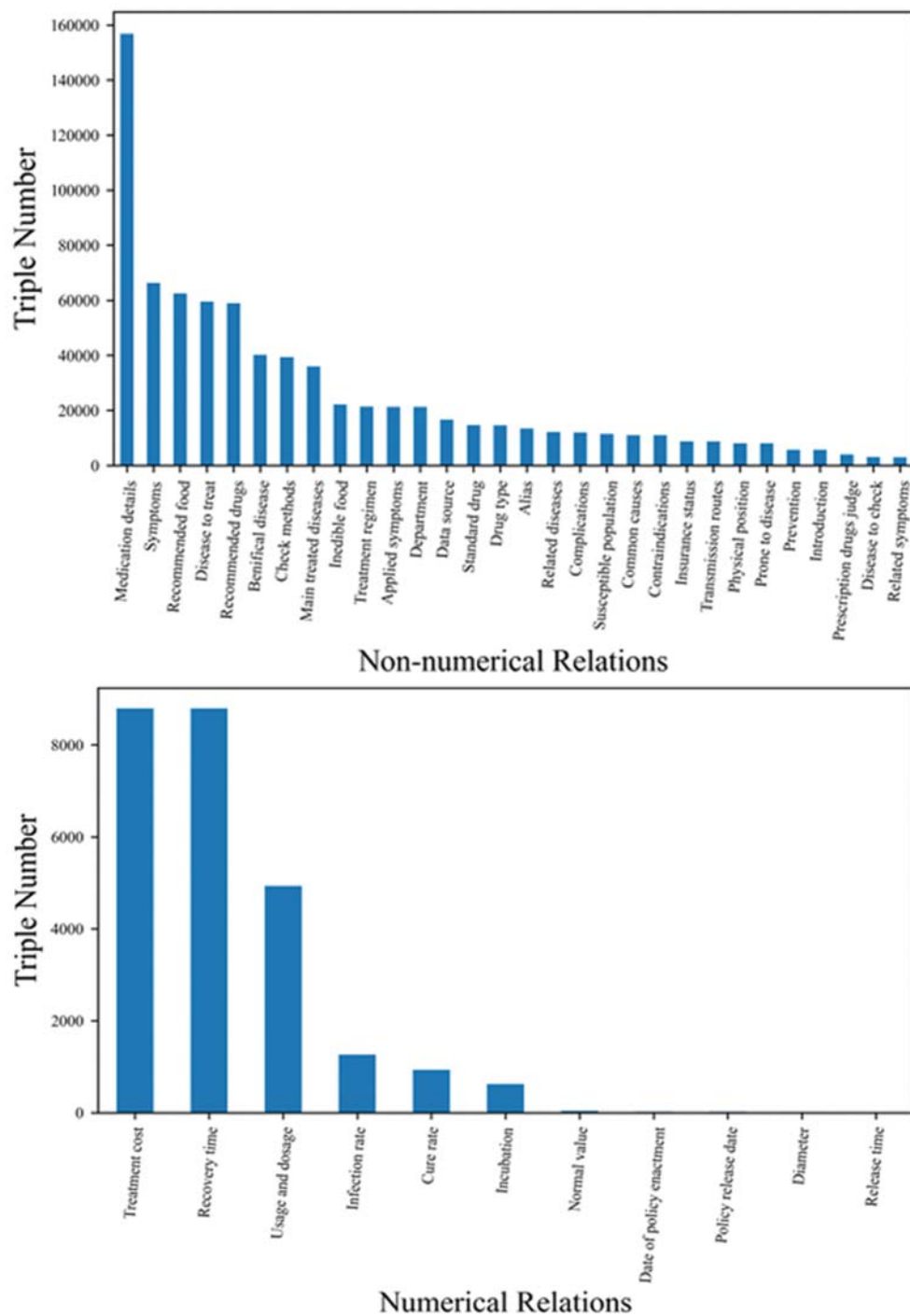


Figure 3. Distribution of COKG-DATA by data types.

to the question in the dataset that has the same multi-hop relations or is most similar to the user's question to be the final response.

7.2 Sources of Answers

We give the sources of answers with the corresponding URL to help users to trace the context, which also increases the credibility of the system. The answer sources of our system give evidence by offering graph names in selected sub-graphs. Multiple graph names are shown if the user's question covers multiple linked graphs. The example can be seen in Figure 4.

7.3 Use Feedback

We design thumbs-up and thumbs-down buttons to encourage users to provide feedback, which will be used to improve the COKG-QA model. When users give positive feedback, the system will randomly generate a thank you sentence. When users thumb down, a bubble will pop up and three options are displayed for users: Incorrect answer, incomplete answer and customized opinions. The custom options provide space for users to flexibly come out with suggestions and further benefit to improve the effectiveness of the QA system.

7.4 Ease of Use

Many medical terms are uncommon or difficult to remember for users, such as disease names and treatments. The automatic input prompt function is significant and practical to improve the usability of the system. Our system supports autocompletion in many scenarios. For example, users can just use a single word, pinyin, first letters of multiple words, or even fuzzy search. Tips in the input box can expand the focus of users' queries to help complete questions that users want to ask as shown in Figure 5. Besides, our system can also recommend questions relevant to the topic entity, which allows users to explore more about the original question.

8. CONCLUSIONS

In this paper, we introduce a multi-hop KGQA method named COKG-QA to develop a QA system for COVID-19 consulting services and meet people's tailored medical information needs. Multi-hop KGQA techniques have attracted increasing attention of researchers for the ability to handle complex multi-hop questions and reasoning. We extend the state-of-the-art method EmbedKGQA by adding Embedding Projection and Schema-Aware Module in this paper. EmbedKGQA represents knowledge graph embedding based on ComplEx and represents questions using RoBERTa. Although it is reasonable and convenient to represent different specified embeddings, these representations come from heterogeneous vector spaces which will influence the optimal performance. We adapt the important features of questions and knowledge embeddings from different spaces into a common semantic one by adopting an embedding projection mechanism. What's more, current KGQA methods ignore the schema implication for entity representation.

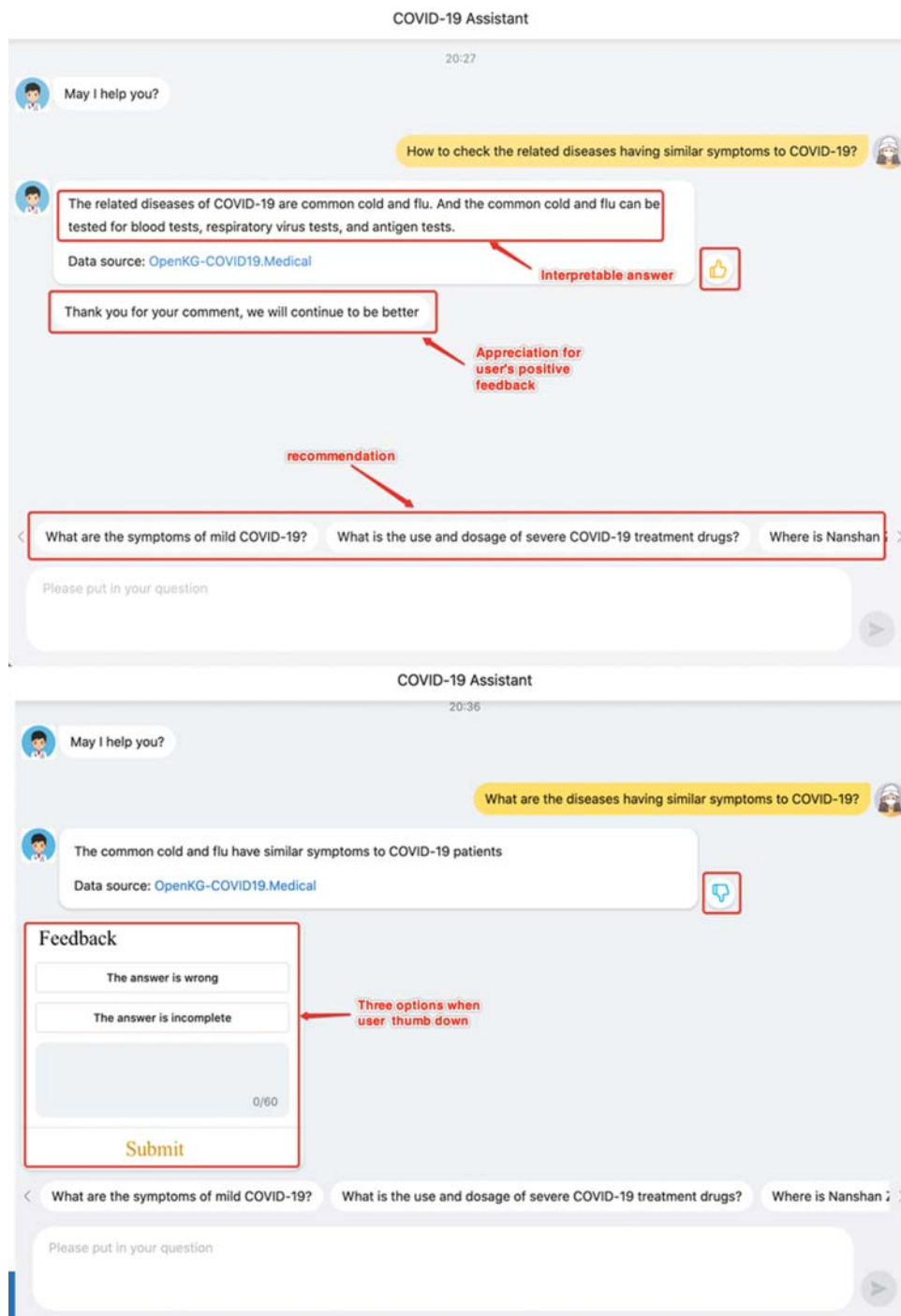


Figure 4. User-friendly functions of our QA system.

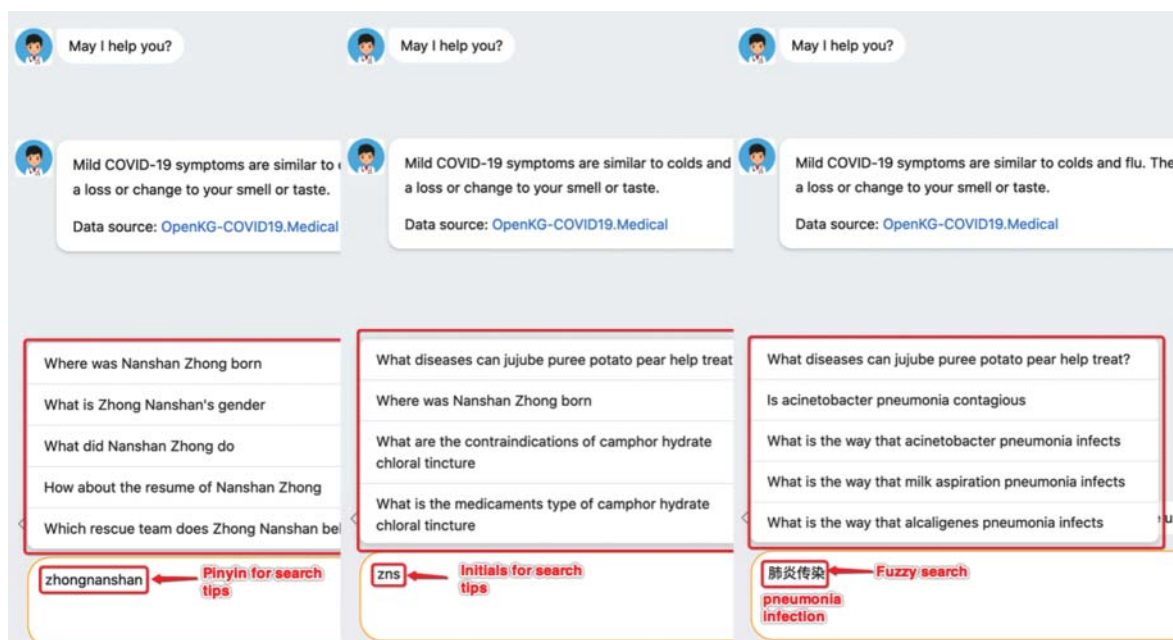


Figure 5. Usability of our QA system.

COKG-QA learns entity embeddings by summing their corresponding type information to help search for the right answer entity of specified types. And to ensure superior performance, we also add a Topic-Entity-Aware Filter to select the answer from the topic entity's neighbor entities in the 3-hop relation range. Furthermore, we publish a large multi-hop Chinese COVID-19 KGQA dataset COKG-DATA based on the open license of CC BY SA to provide a comprehensive knowledge foundation for COKG-QA. Extensive experiment results showed that COKG-QA is robust as a QA engine and can further generalize to new fields. Based on COKG-QA, we also develop a user-friendly interactive application. The application can generate interpretable answers and is easy to use with functions of smart tips and recommendations.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the ScienceDB repository: <https://doi.org/10.57760/sciencedb.02062>. To reuse the data, please cite the data as: Du, H.F., et al.: COKG-QA: Multi-hop question answering over COVID-19 knowledge graphs. Data Intelligence 4(3), 2022. doi: <https://doi.org/10.57760/sciencedb.02062>.

ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Funds for the Central Universities with grant Nos. 22120220069 and the National Nature Science Foundation of China with Grant No. 62176185 and was supported in part by the Shanghai Artificial Intelligence Innovation and Development Fund grant 2020-RGZN-02026.

AUTHOR CONTRIBUTIONS

Du H.F. (duhuifang@tongji.edu.cn), Wang H.F. (carter.whfcarter@gmail.com) designed the model architecture. Le Z.W. (20210240064@fudan.edu.cn) participated in the discussion about the task definition of the project and lead the implementation of the collection of COKG-DATA and the training of COKG-QA with Du H.F.. Du H.F. also developed the Web application. Chen Y.W. (chenyunwen@datagrand.com) and Yu J. (yujing@datagrand.com) added the contrast experiments and made result analysis. All the authors have made valuable contributions in writing and revising the manuscript.

REFERENCES

- [1] Zhang, Y., et al.: WULAI-QA: Web understanding and learning with AI towards document-based question answering against COVID-19. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 898–901 (2021)
- [2] Su, D., et al.: CAiRE-COVID: A question answering and query-focused multi-document summarization system for covid-19 scholarly information management. In Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP (2020)
- [3] Moller, T., et al.: COVID-QA: A question answering dataset for COVID-19. In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL (2020)
- [4] Lee, J., et al.: Answering questions on COVID-19 in real-time. In Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP (2020)
- [5] Ding, K., et al.: Research on question answering system for COVID-19 based on knowledge graph. In 40th Chinese Control Conference, pp. 4659–4664 (2021)
- [6] Michel, F., et al.: Covid-on-the-Web: Knowledge graph and services to advance COVID-19 research. International Semantic Web Conference, pp. 294–310 (2020)
- [7] Sun, Y., et al.: The COVID-19 question answering system based on knowledge graph. In IEEE/ACIS 20th International Fall Conference on Computer and Information Science, pp. 215–220 (2021)
- [8] He, L., et al.: Optimizing automatic question answering system based on disease knowledge graph. Data Analysis and Knowledge Discovery 5(5), 115–26 (2021)
- [9] Sun, H., et al.: Open domain question answering using early fusion of knowledge bases and text. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (2018)
- [10] Sun, H., Bedrax-Weiss, T., Cohen, W.: PullNet: Open domain question answering with iterative retrieval on knowledge bases and text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 2380–2390 (2019)

- [11] Saxena, A., Tripathi, A., Talukdar, P.: Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 4498–4507 (2020)
- [12] Domingo-Fernandez, D., et al.: COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *Bioinformatics* 37(9), 1332–4 (2021)
- [13] Zhang, P., et al.: Toward a coronavirus knowledge graph. *Genes* 12(7), 998 (2021)
- [14] Reese, J.T., et al.: KG-COVID-19: A framework to produce customized knowledge graphs for COVID-19 response.
- [15] Wang, H., et al.: Construction of A Linked Dataset of COVID-19 Knowledge Graphs: Development and Applications. *JMIR Medical Informatics* 26(04), 37215 (forthcoming/in press) (2022)
- [16] Ding, K., et al.: Research on question answering system for COVID-19 based on knowledge graph. In 40th Chinese Control Conference, pp. 4659–4664 (2021)
- [17] Pei, Z., et al.: A general framework for Chinese domain knowledge graph question answering based on TransE. In *Journal of Physics: Conference Series* 1693(1), 012136 (2020)
- [18] Bordes, A., et al.: Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013)
- [19] Du, H., et al.: Progress, challenges and research trends of reasoning in multi-hop knowledge graph based question answering. *Big Data Research* 7(3), 2021026 (2021)
- [20] Hamilton, W.L., et al.: Embedding logical queries on knowledge graphs. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 2030–2041 (2018)
- [21] Ren, H., Hu, W., Leskovec, J.: Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In *International Conference on Learning Representations* (2018)
- [22] Ren, H., Leskovec, J.: Beta embeddings for multi-hop logical reasoning in knowledge graphs. *Advances in Neural Information Processing Systems*, p. 33 (2020)
- [23] Gardner, M., Talukdar, P., Kisiel, B., Mitchell, T.: Improving learning and inference in a large knowledge-base using latent syntactic cues. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 833–838 (2013)
- [24] Xiong, W., Hoang, T., Wang, W.Y.: DeepPath: A reinforcement learning method for knowledge graph reasoning. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 564–573 (2017)
- [25] He, G., et al.: Improving multi-hop knowledge base question answering by learning intermediate supervision signals. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 553–561 (2021)
- [26] Meilicke, C., Chekol, M.W., Ruffinelli, D., Stuckenschmidt.: Anytime bottom-up rule learning for knowledge graph completion. *International Joint Conference on Artificial Intelligence*, pp. 3137–3143 (2019)
- [27] Lin, X.V., Xiong, C., Socher, R., Stuckenschmidt.: Multi-hop knowledge graph reasoning with reward shaping. *United States Patent Application* 16(051), 309 (2019)
- [28] Shi, J., et al.: TransferNet: An Effective and Transparent Framework for Multi-hop Question Answering over Relation Graph. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 4149–4158 (2021)
- [29] Trouillon, T., et al.: Complex embeddings for simple link prediction. *Proceedings of the 33rd International Conference on International Conference on Machine Learning* 48, 2071–2080 (2016)
- [30] Dong, L., Wei, F., Zhou, M., Xu, K.: Question answering over Freebase with multi-column convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Volume 1: Long Papers, pp. 260–269 (2015)

- [31] Bordes, A., Weston, J., Usunier, N.: Open question answering with weakly supervised embedding models. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 165–180 (2014)
- [32] Liu, Y., et al.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
- [33] Zhang, C.L., et al.: In defense of fully connected layers in visual representation transfer. In Pacific Rim Conference on Multimedia, pp. 807–817 (2017)
- [34] Wang, P., Zhou, J., Liu, Y., Zhou, X.: TransET: Knowledge graph embedding with entity types. Electronics 10(12), 1407 (2021)
- [35] Moon, C., Jones, P., Samatova, N.F.: Learning entity type embeddings for knowledge graph completion. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 2215–2218 (2017)
- [36] Saxton, D., Grefenstette, E., Hill, F., Kohli, P.: Analysing mathematical reasoning abilities of neural models. In International Conference on Learning Representations (2018)
- [37] Wang, Z., et al.: Tackling long-tailed relations and uncommon entities in knowledge graph completion. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 250–260 (2019)
- [38] Yamada, I., et al.: LUKE: Deep contextualized entity representations with entity-aware self-attention. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 6442–6454 (2020)
- [39] Vtyurina, A., Savenkov, D., Agichtein, E., Clarke, C.L.: Exploring conversational search with humans, assistants, and wizards. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2187–2193 (2017)
- [40] Podgorny, I., Khaburzaniya, Y., Geisler, J.: Conversational agents and community question answering. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (2019)
- [41] Zhang, Y., et al.: Variational reasoning for question answering with knowledge graph. In Thirty-Second AAAI Conference on Artificial Intelligence (2018)
- [42] Yih, W.T., et al.: The value of semantic parse labeling for knowledge base question answering. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 201–206 (2016)
- [43] Talmor, A., Berant, J.: The Web as a Knowledge-Base for Answering Complex Questions. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 641–651 (2018)

APPENDICES

A. Details of COKG-DATA

A.1 Overview of COKG-DATA

We elaborately select seven sub-graphs that contain topics people are more concerned about during the COVID-19 epidemic. The specific graphs selected by COKG-DATA are demonstrated as follows.

- The encyclopedia KG gives us a general understanding of SARS-CoV-2 and COVID-19, and relevant viruses and diseases information.
- The prevention KG provides prevention guidance published by the government for individuals, organizations in different places.
- The goods KG is expanded around materials supply status during the epidemic, covering daily protective equipment, medical devices, and drugs.
- The medical KG and the health KG are complementary to exploit COVID-19 related knowledge about various diseases, drugs, symptoms, examination methods, and hospitals.
- The epidemiology KG employs the general techniques of epidemiology to study the distribution of diseases and influencing factors, exploring the causes of disease, clarifying the laws of epidemics for controlling and eradicating diseases effectively.
- The character KG records concepts such as characters, battles, achievements for the pandemic, articles, resumes of heroes, etc.

A.2 Data Curation

Data cleaning. To ensure the quality of the QA dataset, we have cleaned some bad cases in OpenKG-COVID19 and removed triples that are not practical for QA: (1) some triples contain empty string, punctuation entities, or useless numbers; (2) some triples are weird to compose natural questions, e.g., {Doctors of Xinhua hospital, work in, Xinhua hospital} (3) the head entity is same with the tail entity in some triples, such as triples with “alias” relation. We filter out these bad triples described above and remove them. In addition, relation patterns including symmetry and inversion exist in OpenKG-COVID19. We extend triples for these relation patterns of OpenKG-COVID19. After data cleaning and relation extension, the knowledge graph dataset contains 112,246 entities, 209 relations, and 787,056 triples.

Multi-hop Questions Collection. We leverage fact triples in the selected sub-graphs of OpenKG-COVID19 as single-hop data. Further, we manually design 47 relations for 2-hop questions and 23 relations for 3-hop questions, in which the combined relations must be reasonable and natural. Specifically, the range of the front relation must be the same with the domain of the back relation in a 2-hop relation. For example, the range of “selected drug” relation is “drug” which must be consistent with the domain of “usage and dosage” in the 2-hop relation “Selected drug Usage and dosage”. The same rule applies to the 3-hop relations collection process. Similar to multi-hop dataset MetaQA [41], we employ neural translation models in

Helsinki-NLP Opus-MT project[®] to introduce more diverse and natural statements with the same meaning. Opus-mt-zh-en model is leveraged to translate sentences from Chinese to English, and then opus-mt-zh-en is used to translate back to Chinese. Furthermore, to create a large-scale unified knowledge base from the top level, entity alignment and relation alignment have been completed to eliminate inconsistency problems.

[®] <https://github.com/Helsinki-NLP/Opus-MT>, a project offers tools and resources for open translation services

AUTHOR BIOGRAPHY



Huifang Du is a Ph.D student in the college of Design and Innovation, Tongji University. Her research areas are knowledge graph and question answering. ORCID: 0000-0002-5241-7620



Zhongwen Le is a graduate student in the School of Computer Science, Fudan University. His research areas are knowledge representation and question answering. ORCID: 0000-0001-9364-1218



Haofen Wang has long served as Chief Technology Officer in a firstline artificial intelligence company and has rich experience in AI RD management. He is one of the founders of OpenKG, the world's largest Chinese open knowledge graph alliance. He is responsible for participating in a number of provincial and ministerial AI-related projects, and has published more than 100 high-level papers in the AI field, which have been cited more than 2,300 times and the H-Index has reached 23. He built the world's first interactive virtual idol—"Amber·Xuyan"; the intelligent customer service robot he built has served more than 1 billion users. Currently, he is the Deputy Director of the Terminology Committee of China Computer Society, the Deputy Secretary General of the Language and Knowledge Computing Committee of Chinese Information Society and distinguished research fellow in College of Design Innovation, Tongji University. ORCID: 0000-0003-3018-3824



Yunwen Chen received the Ph.D. degree from Fudan University, Shanghai, China. He is the founder and the CEO of DataGrand Inc., Shanghai, a leading AI company in China. He had been the Chief Data Officer of Shanda, Inc., Burlington, IA, USA, the Senior Director of Tencent, Inc., Shenzhen, China, and a Researcher of Baidu, Inc., Beijing, China. He has 32 patents and several academic publications. His current research interests include data mining, natural language processing, search and recommend systems, and knowledge graphs. Dr. Chen was a recipient of the Distinguished Graduate Student in 2008. He is a Senior Member of the CCF and a member of the ACM. ORCID: 0000-0003-4513-9439



Jing Yu received the M.S. degree from Tongji University, Shanghai, China. He is currently the co-founder and the department manager of DataGrand Inc. in China. His research interests include search and recommender systems, knowledge graph, natural language processing and machine learning etc. He has tens of patents and several academic publications in the artificial intelligence. He is a Senior Member of the CCF and a member of the ACM. ORCID: 0000-0003-3655-2811